1460

A reprint from April, 1970
Vol. 19, No. 4
computers and automation

"THE EMPTY COLUMN" REVISITED

William J. Wiswesser
Fort Detrick
Frederick, Md. 21701

A Chemical Notation that Appeared with Computer Languages in 1950

8

"THE EMPTY COLUMN" REVISITED

William J. Wiswesser
Fort Detrick
Frederick, Md. 21701

A Chemical Notation that Appeared with Computer Languages in 1950

"These are just starting examples of computer benefits that the chemical world will enjoy when more manpower, money and talented attention is devoted to this 20-year-old chemical notation with the empty columns."

William J. Wiswesser, a research chemist at Fort Detrick, Frederick, Md., probably is best known as the inventor of the Wiswesser Line Notation (WLN), which "The Empty Column" parable introduced 20 years ago. He is a native Pennsylvanian, graduated from Lehigh University in 1936, later taught chemical engineering courses at Cooper Union, and probably created the WLN as a hybrid of long-rooted interests in atomic art, molecular structure, history of chemistry, and information theory.

The parable about a "New Notation" of Long Ago (***Computers and Automation,*** January 1970, page 16) has a significance that was not fully appreciated when it was written twenty years ago – that this imagined rejection of Arabic numerals by users of Roman numerals ***may have occurred many times*** during the past two thousand years! Medieval merchants were jailed if they were caught manipulating "those heathen signs and symbols". The battle lasted for some 300 years, because official examiners – like the Roman in the parable – just did not see how the positional Arabic numeration profoundly simplified all mathematical operations.

Martin Gardner gave the following fascinating background details on this mathematical blindness in the January 1970 issue of ***Scientific American*** (pages 124-125):

> For more than 15 centuries the Greeks and Romans and then Europeans of the Middle Ages and early Renaissance calculated on devices with authentic place-value systems in which zero was represented by an empty line or groove or by an empty position on the line or groove. Yet when these same people calculated without mechanical aids, they used clumsy notational systems lacking both place values and zeros. It took a long time [from 1202 to the 16th century] . . . to realize that in writing numbers efficiently it is necessary to draw a symbol to indicate that a place in the number symbolizes nothing.
>
> . . . In some European countries calculating by 'algorism' actually was forbidden by law, so that it had to be done in secret. There was opposition to it

even in some Arabic countries. Not until paper became plentiful in the 16th century did the new notation finally win out, and soon after that the shapes of the 10 digits became standardized because of printing.

The corresponding need today for simplified chemical descriptions should become obvious with just three relatively simple statements, but chemists – like all humans – continue to overlook the obvious:

(1) all chemical information has a cosmic common denominator – the sharply defined atom-to-atom structure descriptions;
(2) there are some 4,000,000 such reported structures in the chemical world – needing concise computer descriptions for their efficient retrieval; and
(3) the most frequently used atomic symbols ***and groups*** should be ***single***-mark symbols.

This last point was made 157 years ago by J. J. Berzelius, "the organizer of chemistry" and editor of many pioneering chemical journals. But his point was soon forgotten. Computers can help the chemists far more if the chemists recognize and provide a notation that reflects overall "least effort" (in the ***long***-term view!). Least effort implies being easy to learn, to write, and to remember – easy to use in every man/machine aspect.

Line-Formula Notations

The occasion for writing the "Empty Column" parable was an internationally publicized development – the search for an international chemical notation by a "Commission on Codification, Ciphering, and Punched Card Techniques," established in 1947 by the International Union of Pure and Applied Chemistry (IUPAC). In 1949 the author had been appointed to serve in what then was called the "Punched Card Committee" of the American Chemical Society; he wrote this parable a year later (May 1950) as a needed preface to his proposed standardization of "line-formula" structure descriptions. Chemists had been using "rational formulae" or "line formulas" as ***delineated*** structure descriptions, ever since the age of Structural Chemistry dawned in 1861. All that seemed necessary was a careful standardization for tabulating equipment (and today's computers) of this world-wide, time-tested tradition. The parable was written as a caution to the IUPAC and other examiners that any new notation may have a strange and puzzling appearance at first glance.

Cosmic Identification

Line-formula notations developed in a simple and natural way that most chemistry accounts overlook; so a few explanatory figures and historic examples seem appropriate here. The cosmic identification of a chemical compound is its structural (or constitutional or "rational") formula – a two-dimensional diagram showing how all the atoms in a molecule are connected. Thus the three structure diagrams in Figure A not only explain "rationally" what the substances are – they also explain how ethyl acetate can be hydrolyzed (split apart by the addition of H-O-H or water and suitable catalyst) to ethyl alcohol and acetic acid, or how the alcohol and acid combine to form the ***ester*** with a suitable dehydrating agent.

The corresponding "new" notations (introduced with the parable 20 years ago) are given under the names in Figure A. These notations reflect a natural reduction in writing effort that started almost as soon as structure diagrams appeared. Thus within a brief seven-year period (1861-1868), simpler and more compact linear expressions replaced the two-dimensional diagrams in journal discussions: the 2-carbon "ethyl" chain was contracted to $CH_3.CH_2$- or CH_3CH_2- or C_2H_5- or simply ***Et*** marks. The corresponding "acetyl" group was simplified to $CH_3.CO$.- or CH_3CO- or simply ***Ac*** marks. Thus to this day ethyl alcohol is frequently symbolized as EtOH, acetic acid as AcOH, and ethyl acetate as EtOAc. The corresponding "new" notations Q2, QV1, and 2QV1 give even more concise descriptions, with simpler typography and more logical (language-free) sets of symbols.

Comparing Old and New

Table 1 compares these names, old line-formulas, and new notations with those of other related and important

Table 1. UNBRANCHED OPEN-CHAIN COMPOUNDS

NAME	OLD LINE-FORMULA	NEW NOTATION
acetone	CH_3-CO.CH_3	1V1
ethyl ether	C_2H_5-O-C_2H_5	2O2
ethyl acetate	C_2H_5-O-CO.CH_3	2OV1
butyl acetate	$CH_3CH_2CH_2CH_2$-O-CO.CH_3	4OV1
ethyl alcohol	CH_3CH_2-OH	Q2
acetic acid	CH_3-CO.OH	QV1
carbonic acid	HO-CO.OH	QVQ
ethylamine	CH_3CH_2-NH_2	Z2
acetamide	CH_3-CO.NH_2	ZV1
urea	NH_2-CO.NH_2	ZVZ

Note: The period in the CO-groups denotes the end of a doubly-bonded or :O side group, distinguishing this from an -O- link.

compounds. The structure diagrams for the hydrocarbon fragments are like those shown in Figure A. An amateur code-breaker can see at a glance that analogous things have analogous notation symbols: ***numerals*** denote the number of carbon atoms in the hydrocarbon chains, and ***letters*** denote "functional" groups that characterize the chemical types. For example, ***alcohols*** have the lone -OH or Q-terminal, ***ethers*** the lone -O- link, and ***ketones*** the lone -CO.- or -V- link; ***acids*** have the -CO.OH or -VQ combination, and ***esters*** the -O-CO.- or -OV- combination. (The period in -CO.- denotes the ***end*** of the :O ***side*** group, distinguishing it clearly from the connecting -O- link.)

Nitrogen analogs of alcohols and acids also have notations that show more direct similarities than the corresponding (unspaced ***amine*** and ***amide***) names. The appropriate pairs in Table 1 are those in which the terminal -OH or Q-group is replaced by a -NH_2 or Z-group: Q2 and Z2, QV1 and ZV1, QVQ and ZVZ.

ETHYL ALCOHOL — ACETIC ACID — ETHYL ACETATE

Q2 — QV1 — 2QV1

Figure A.

Branched Structure

The first branched structure in Figure B, copied from an 1866 report, shows how naturally the line-formula convention arose as a one-dimensional printing simplification of two-dimensional structure diagrams. At that time the "carbon skeleton" usually was drawn vertically, like the human

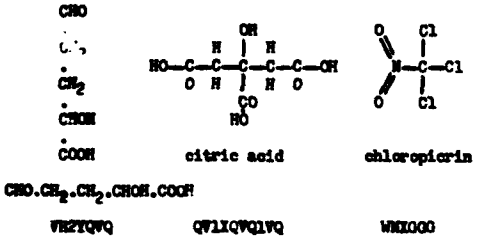

Figure B.

skeleton, but with ***all*** of the "appendages" extending to the right (and in more compact groups than those shown in Figure A). Thus the line-formula delineation of these compacted groups is simply a television-like scanning of the two-dimensional diagram – left to right and top to bottom. This illustrated notation introduces two new features: a terminal VH-group for the top aldehyde or CHO-group, and a Y-symbol for the Y-branched or ternary carbon (attached to three atoms other than hydrogen). This ***branching*** distinction is a very important "connection table" specification. The ***linking*** -CH_2CH_2-group is denoted simply as a 2-carbon chain, without the extra H-atom that the corresponding ***terminal*** chain must have.

Citric acid, the second example in Figure B, illustrates a typical partial compacting of the pictured groups; the reduced cluttering of lines emphasizes the distinct X-branching nature of the central carbon atom; hence the X-symbol denotes a quaternary carbon (attached to four atoms other than hydrogen).

Chloropicrin, the third example in Figure B, also illustrates an X-branched carbon and two other new features. (1) a ***single*** G-mark "fusion" of the Cl symbol for the very frequently cited chlorine atoms; and (2) a branched dioxygen group, important enough to be denoted by a single-letter ***W*** (its "double-U" name alludes to the two double-bond connections seen in most branched dioxygen structures).

The R-mark

Three graphically distinct kinds of benzene derivatives are illustrated in Figure C. All have a characteristic regular-hexagonal C_6-ring that is more prominent in chemical

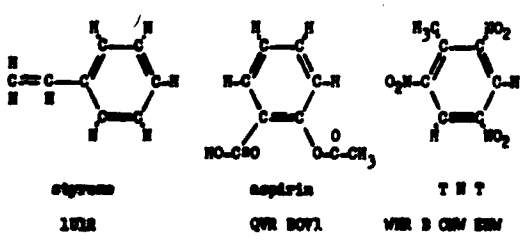

Figure C.

catalogs than all other rings combined. Accordingly this ring is denoted most efficiently as a ***single*** mark – the letter ***R*** (for Ring) – and subordinated to all other atomic-group symbols because of its superprominence. This R-mark saves more writing effort than any other notation mark (reflecting traditional abbreviations ***Ph*** or the "phi" sign ϕ for the ***phenyl*** or C_6H_5-group); it also eliminates the graphical need to show the ring-forming connections as alternating or "resonating" single and double bonds, often called ***aromatic*** bonds to distinguish them from the quite different open-chain double bonds.

Styrene, the first example in Figure C, illustrates the open-chain kind of double bond, an ***unsaturation*** – hence denoted with the letter ***U***. These groups are so active that they will spontaneously link together, forming the ***saturated*** chains of polystyrene, with a C_6H_5 or phenyl side group on every other chain atom. Many other phenyl or C_6H_5-derivatives (with only one replaced H-atom), like styrene, have structurally unrevealing names and pictorially direct notations. A few of these many examples are listed in Table 2.

Table 2. COMMON PHENYL DERIVATIVES

NAME	OLD LINE-FORMULA	NEW NOTATION
anisole	C_6H_5-O-CH_3	1OR
toluene	C_6H_5-CH_3	1R
styrene	C_6H_5-CH:CH_2	1U1R
phenol	C_6H_5-OH	QR
benzoic acid	C_6H_5-CO.OH	QVR
nitrobenzene	C_6H_5-NO_2	WNR
aniline	C_6H_5-NH_2	ZR

Note: The C_6H_5-ring fragment is frequently denoted as Ph or Ø (phi). In the new notation, the ZERO mark is slashed as a Ø mark.

Aspirin, the second example in Figure C, appropriately shows the "empty column" solution to what is a real headache in many other chemical notations: the need for a logically distinct set of symbols to locate ring positions. In 1866 Kekulé used ***lower case letters*** for this purpose, so in 1950 his meaning was put into "Teletype" equivalents ***by prefixing each locant letter with a blank space.***

TNT, the third example, illustrates how this spaced locant alone suffices when the located group is the commonplace ***methyl*** group or unit-carbon chain.

Space as a Mathematical Operator

The "Empty Column" thus seemed an appropriate title for the parable because a corresponding "empty" or ***blank*** space is an essential and unique part of the notation that it prefaced: this SPACE serves as a mathematical operator or shift key to ***convey lower case meaning to the letter that follows,*** and all such ***LO***wer ***CA***se let***TE***rs LOCATE ring positions. This spaced "locant" also begins a new unit of information, mentally translating to mean "and at this ring location the following atomic group is attached." Thus in addition to the gain of a doubled keyboard without a penny of cost, the heavily used spaces facilitate manual reading, like the spaces between words. Similarly spaced ***numerals*** also give them distinct meaning as ***multipliers*** of the preceding string of symbols; these operate like a "Polish string notation" in omitting the need for quantity-enclosing

marks, which were not available in 1950-vintage tabulating equipment.

Other notation designers overlooked this obviously profitable use of a "blank space" character, but that is not surprising to historians: the Greeks and Romans, for all their intelligence,

> ranntheirwordstogetherlikethisbecausetheydidnot realizethat SPACES greatly facilitated the reading thereof! This spacing of words also was a medieval discovery.

In 1950 there were no punctuation marks available other than the ampersand, which has served well ever since then to *end* side groups other than the few that are strictly terminal by definition (like the illustrated G, H and Q marks). Notations for all ring structures other than the C_6-hexagon of benzene ideally were enclosed in parentheses, and the 1950 letter-substitutes for *carbo*cyclic ring notations were inspired from an 1866 diagram. In that year Emil Erlenmeyer (the flask man) tried to explain the two-ring structure of naphthalene with the diagram shown in Figure D. His L-shaped and J-shaped marks indicated a

NAPHTHALENE (Erlenmeyer, 1866) VITAMIN B_6

L66J T6NJ B CQ D1Q E1Q

Figure D.

connecting line between those carbon atoms; this suggested the use of L. . .J marks to "enclose" carbocyclic (including a*L*icyclic) ring-descriptions, and T. . .J to enclose he*T*erocyclic equivalents. Rings in general can have so many topological complications that it is not possible to summarize other details here. Vitamins B_6 in Figure D is a *hetero*cyclic compound of average complexity.

"Connection Table" Specifications

The first rule of this "empty column" chemical notation is to cite chains of atomic groups in end-to-end connecting order, following the line-formula tradition. The "least effort" gain is that no search has to be made for some arbitrarily preferred "central component," as in the IUPAC notation, and no related "assembly instructions" are needed for the pictorially direct attachments. The gain in minimizing "connection table" specifications seems so obvious that one wonders why others had not applied this same gain in complicated ring systems, where this least-effort notation follows a longest-possible path of connections. This maximized path thereby minimizes ideal ring descriptions to a simple recitation of the *non*consecutive links.

The second rule also is so simple and obvious that it was overlooked until this line-formula notation appeared in 1950: Resolve all otherwise equal alternatives by the simple alpha-numeric order of the notation symbols. Long afterward, this proved to be the simplest thing a computer could do: compare "equals" until a higher or lower resolution is reached! Even here, intellectual complications have become rooted; thus in 1950 the notation followed the seemingly natural Hollerith-sorting sequence of *numbers* before *letters*. (We could not imagine anyone counting his peanuts as A, B, C, and then when he ran out of letters, going to 1, 2, 3!) The 1950 terminology defined the letters as having *higher* rank than the numerals, just as the value of 9 is higher than that of 1. The notation's rule 2 specified a *descending* citing order – letters before numbers, because in open-chain structures the letters feature the characteristic chemical functions like acid, alcohol and aldehyde, these determine the properties and uses, whereas the *numbers* denote the *number of carbon atoms* in the relatively inactive paraffin chains (*Par affinis* means low affinity or low activity). Thus rule 2 tends to bring together chemically similar things like open-chain alcohols in simple, alphabetically arranged lists like those in Tables 1 and 2.

Pope Paul described this "least effort" aim when he advised "Avoid complicating simple things; strive to simplify complicated things."

The Character Set

The "program language" of this *chemist*-oriented notation is best illustrated, not with more recited rules, but with a summarizing review of the basic descriptive tools – the character set. If these are well chosen, and cited in pictorially direct connecting order, the rules for handling them almost come naturally.

Berzelius, as previously noted, gave the first long-overlooked requirement for citing chemical structures with least effort: the most frequently cited atomic groups should have *single* marks. Thus in 1813 he established nine perfect choices for the very frequently cited *non*metallic atoms of boron, carbon, fluorine, hydrogen, iodine, nitrogen, oxygen, phosphorus, and sulfur. His apt recommendations can be remembered as a tic-tac-toe that appropriately "begins with BC," "has an *I* in the middle," and appropriately "ends with a PS." (See first part of Figure E.)

Figure E.

Bromine was not yet discovered when Berzelius assigned B for boron. Today it is extracted from the sea in ton-a-day plants (to make lead-scavenging gasoline additives like ethylene dibromide), so the ideal *single*-letter symbol is "extracted" from the front part of the *Br* symbol. An equally obvious clue was overlooked until some ten years ago, when a Syracuse University student showed the lecture audience that the hinted *E* can be extracted directly from s*E*a!

Chlorine was first known by an appropriately frightening appellation as "dephlogisticated muriatic acid gas"; so Berzelius aptly assigned a single letter *M* for the muriatic radical in his first (1813) list of atomic symbols. To this day the *Cl* replacement continues to give trouble in letter-number ambiguities, so these are fused into a *single*-letter *G*, the 7th letter of the alphabet for the leading atom in the 7th Group of the Periodic System. This choice is triply appropriate because *G* and *E* stand next to each other in the word halo*GE*n as well as in the Periodic Table. The symbols F, H and I combine with these to form *an alphabetically closed set*, with obvious indexing advantages.

Lengthening

Berzelius analyzed the importance of symbol selections so well that no new single-letter symbols need to be assigned for high-frequency structural atoms, other than the above *E* and *G* for bromine and chlorine atoms. However, he and his followers overlooked an obvious gain in his original intent to give *all* metallic atoms *two*-letter symbols, like his original *PO* for potassium; metallic atoms are cited

much less frequently than nonmetallic atoms, and there are far more kinds of them. Trends in "least effort" usage gave overlooked clues like the ***lengthening*** (for easier recognition) of ***L*** to ***Li*** and ***R*** to ***Rh*** for the rarely used lithium and rhodium symbols. A late 19th century Harvard textbook of chemistry showed a more helpful ***Ur*** instead of ***U*** for uranium, ***Va*** instead of ***V*** for vanadium, and ***Wo*** instead of ***W*** for wolfram or tungsten. A more recent (1921) textbook of inorganic chemistry from M.I.T., and other reference books of that period, showed ***Yt*** instead of ***Y*** for the really rare yttrium in Periodic Tables. Chemistry students would welcome the simplification that ***all*** metals have ***two***-letter symbols, as in this notation. The generalization extended to the equally rare noble gases, for the 1954 notation manual used ***Ar*** to denote argon before IUPAC made this an official international atomic symbol.

Computer Restrictions

What happens when these two-letter symbols must be written with the "Teletype" and computer restrictions of strictly upper-case letters? Here another aid to recognition was overlooked and insufficiently generalized in the original 1954 manual: ***All*** two-letter symbols now are set off in hyphens. Then the computer chemistry can wax poetic and show that -AR- pairs with -KR- in Periodic Group 8, -KA- with -NA- (after Latin ***kalium*** and ***natrium***) in Group 1, -VA- with -TA- in Group 5, and two for good measure in Group 6: -UR- with -CR- and -WO- as a "spitting image" of -MO- Rare -YT- matches the rare earth -YB- in Group 3

The hyphenation intensifies recognition in printed lists, and the two-letter standardization releases six precious single letters for nonmetallic structural groups, most of them cited more frequently than the previously introduced ***E*** and ***G***.

Astronauts as well as aquanauts now behold the beauty of our water-covered blue earth. The OH-group always had great prominence in AQUEOUS chemistry, and now it has cosmic prominence as a free OH radical in outer space. The obvious single-letter choice for this very important group is extracted from pure or polluted AQUA, and this old letter ***Q*** can be well remembered as an O-atom with an H-tail (Figure E, center).

(Old radioman-practice slashes the ***zero***, not the frequently used letter O).

Nitrogen chemistry parallels oxygen chemistry in many ways, but this can be shown more refreshingly with a programming aim to have the important NH-group match the OH-group in retrieval sharpness. The notation symbol for this linking or Mid-aMino NH-group is carefully selected from the middle of the alphabet: the nitrogen counterpart for ***Q*** is the letter ***M***, an N-atom with an H-prop (Figure E).

Computer Retrieval

Carbon, of course, is the characteristic element of the ***organic*** compounds that comprise some 94% of the 4,000,000 reported chemicals; and carbon atoms are found among these structures far more frequently than any other atoms excepting the stellar-wide hydrogen atoms; more frequently, in fact, than all others combined. Thus good computer retrieval requires distinctive single-letter symbols for the different kinds of combined carbon. The obviously best choice for an X-branched carbon atom in open-chain structures is the letter X (denoting a quaternary carbon, or one connected to ***four*** non-hydrogen atoms). Its quaternary nitrogen parallel is denoted with the letter K, the characteristic feature of "kwat" and "kationik" salts. The Y-branched CH-group likewise is best denoted with the letter Y (a carbon or CH-group attached to ***three*** non-hydrogen atoms). A related *V*ery common di*V*alent connecti*V*e, the -CO.-group, is denoted with the letter V (first part of Figure F).

Roman stone-masons made a "least-effort" V-cut for the vowel U. (Its F-related meaning was a later medieval addition to the alphabet.) Thus the V-group has within itself an etomologically related U-mark, elsewhere used for the *U*nsaturating double-bond link. This notation gives considerable ***freedom from chemical bondage***, because the unsaturating symbol ***U*** is used only when it is necessary to show the corresponding physical removal of H-atoms from the connected carbon group (as in the previously illustrated and listed styrene, 1U1R). A third related letter ***W*** was chosen to denote the ***branched dioxygen*** part of nitro and analogous O_2-groups, because it literally whispers its embedded ***double-U*** bonding pattern! (Figure F). The letter W also was a medieval addition to the English alphabet, designed to represent the "UU" or long "ooo" vowel sound, hence its double-u name.

Figure F.

Since the benzene ring occurs more frequently in structure descriptions than all other rings combined (including benzo-fused rings with the others), the most appropriate remaining letter selection for this *R*esonating, *R*egular-hexagonal *R*ing therefore is the letter ***R***, visualized as in Figure F with two adjacent (or ***ortho***) attachments. The enclosed circle in this diagram is the logical "least-effort" way of showing the "resonating" or alternating double bonds. (The author was circling his benzene rings in this "lazy-boy" manner some 35 years ago, so it is hardly a modern innovation.)

Mnemonic Associations

Fastidious professors may feel deeply annoyed by the mnemonic associations in these single-letter selections, and they are not likely to be "turned" by the last of these "dirty dozen" memorizing irritants: the terminal NH_2-group in this notation is denoted with the terminal letter Z (from aZine and hydraZine), a doubly appropriate selection because it is pictorially the very same as the letter N turned ***on end*** (end of Figure F and end of the program-language remarks!).

Perhaps the best way to emphasize and summarize this "Empty Column" lesson about resisting change is a brief recitation of what other users —thousands of miles away— have done with this chemical notation in spite of its officially unrecognized status. About ten years ago the users simplified its identification; people have endless difficulty with this three-syllable, nine-letter WISWESSER word (a lifelong lesson to its bearer), so they speak of the Wiswesser Line Notation simply as the WLN.

This WLN now has an "authorized manual," voluntarily written by Elbert G. Smith (Professor of Chemistry at Mills College in Oakland, California) and published by McGraw-Hill in 1968 — after eight years of rule revisions and user-tested improvements. All royalties from this book go to a Chemical Notation Association, organized in 1965 "(1) to promote and conduct research in the field of chemical notation systems and to advance the development and application of these systems; (2) to educate chemists in the uses and advantages of these systems; and (3) to act as an official adjudicating body to determine and control the

standard rules of any chemical notation system entrusted to this Association for this purpose by its authors, inventors, and developers." The 70-some members of this Association in the United States, the United Kingdom, and France are still concerned with only one notation: the WLN.

The appendix contains a partial list of organizations that have put an investing interest in WLN, as evidenced by publications. Programs based on this standardized line-formula notation now have daily usage in IBM 360 or 1130, Burroughs 5500, Honeywell 200 or 400, CDC 3150, GE 635, PDP-10 and other computers in chemical information centers throughout the world.

Dow's CHECKER Program

One pioneering program, known as Dow's CHECKER program, calculates a molecular formula from the notation and compares this with manually calculated input formula; notation errors are about 2% and formula errors are the same order of magnitude – around 2%.

"WLN-permuting programs" identify another series of routines for IBM, Burroughs, UNIVAC, Honeywell, and GE computers. These programs "permute" or rotate the notation records such that the repeatedly offset atomic symbols form a "key-letter-in-context" alphabetized list. Copies of this WLN-permuting routine have passed around at least a half-dozen computer centers in the United States.

Imperial Chemical Industries, Ltd. have CROSSBOW programs that generate three possible outputs from input WLN records: (1) connection tables for fine structure searching, (2) open-ended "fragmentation codes", which are chemically significant structural components that can be printed as WLN symbol clusters and organized into file records, and (3) computer-generated and high-speed printed-composed structure diagrams. Complementing programs elsewhere are now in process to yield computer-composed notations from input tapes of connection tables, or from hand-drawn diagrams made with light-pen communication by a clerk or chemist at the console.

The PATHFINDER Program

The PATHFINDER program, written for Dow's Burroughs 5500 computer, is a very powerful routine that exhaustively checks all trial paths in extremely complicated ring structures, holding the correct lower-valued choice in all comparisons; the final holding is converted to the infallibly correct carbocyclic notation. Its input is our long-overlooked "nonconsecutive links".

Binary "bit screens" can be searched at phenomenal speed, compared with higher-language alternatives that suffer much input/output processing translation. A computer-generated equivalent of the 1950-vintage multi-punched cards makes binary "scratches" for the distinctively spaced or unspaced WLN symbols and yields a ***30-fold*** increase in speed in sophisticated chemical structure searches.

These are just starting examples of computer benefits that the chemical world will enjoy when more manpower, money and talented attention is devoted to this 20-year-old chemical notation with the empty columns.

The "least effort" advantages of the author's proposed "Line-Formula Chemical Notation" were not acknowledged at the decisive meeting by representatives of IUPAC and the ACS Punched Card Committee, held at M.I.T. in August 1951: the IUPAC examiners decided to "give the axe to the line-formula tradition" and favored an unfamiliar departure that has a more complicated set of resolving rules and a much more complicated character set. It has two or three known users in the chemical world today, in spite of a number of official promotional efforts by the IUPAC authorities.

Like the Arabian mathematician in the parable, we can only guess why we failed to interest official examiners at M.I.T. in 1951 – and elsewhere since then. Perhaps the simplest and most obvious solutions to complicated problems are the most easily overlooked. The power of the human brain to deceive itself – even when healthy and free of disabling drugs – must not be underestimated. We submit the comparisons listed in Table 3 of century-old line formulas and their standardized WLN equivalents, for those who wish to see the conservative correspondence with tradition.

Table 3. COMPARISON OF EARLY LINE FORMULAS (1861-1867) WITH WLN

1861-1867 Line Formula	WLN
$C_2H_5.O.C_3H_4O.O.C_2H_5$	2OV2O2
$CH_2CN.CO.Br$	NC1VE
$ClCH_2\text{-}CO_2H$	QV1G
$H_2N\text{-}CH_2\text{-}CH_2\text{-}CO_2H$	Z2VQ
$CHCl_2.CCl_3$	GYGXGGG
$CO.OH\text{-}C(OH)_2\text{-}CO.OH$	QVXQQVQ
$CH_3.CH1.COOH$	QVY1
$CH_3\text{-}CH.OH\text{-}CO.OH$	QYVQ
$C_6H_5\text{-}CH_2\text{-}Br$	E1R
$C_6H_5.CCl_2H$	GYGR
$C_6HBrHBrHNO_2$	WNR CE EE
$C_6H_5.SO_2.OH$	WSQR

Note: The CH_3-groups attached to Y-branched C-atoms are understood by definition of the Y mark (or X).

The last cited report on computer applications of the WLN (44) gives in its appendix some 500 additional examples, all identified by common name; most of them are grouped into 18 sets, sequenced in increasing order of structural complexity. The chronological arrangement of the 71 reference citations in this same report also reflects the "exponential" growth of user interest in the WLN: only twelve references appeared in the first ten years (1950-1959), then ten in the next five years (1960-1964), followed by twelve in two years (1965-1966), seven in 1967, and no less than nineteen in 1968. This is gratifying growth!

We acknowledge the growing signs of user interest in the WLN as a keen appreciation of their interest, and we submit this "excursion in symbol-land" as special thanks to ***Computers and Automation*** for recognition of the parable that was written in 1950 to introduce our "empty column" notation. □

APPENDIX

A partial list of organizations that have expressed an interest in the WLN and published or presented papers on it, is given below. Their reports are keyed to the numbers in the literature references (which also include the earliest citations on the WLN).

J. T. Baker Chemical Co. (2, 3, 26, 43)
Chemical Abstracts Service (9)

Diamond-Shamrock Corporation (13, 14, 26)
Dow Chemical Company (7, 8, 26)
Food & Drug Administration (1)
GAF Corporation (35)
Goodyear Tire & Rubber Co. (10)
Hebrew University (Israel) (19)
Hoffmann-LaRoche, Inc. (26, 34)
Imperial Chemical Industries (21, 22, 26, 27)
Institute for Scientific Information (23, 28)
Eli Lilly and Company (29)
Mills College (E. G. Smith) (31, 32, 33)
Ministry of Defense of Israel (27)
National Bureau of Standards (11)
National Library of Medicine (1, 30)
Olin Mathieson Corporation (18)
G. D. Searle & Co., Inc. (5, 6)
Stanford Research Institute (20, 26)
University of Pennsylvania (24, 26)
University of Sheffield (UK) (25)
U. S. Army, CIDS Program (26)
U. S. Army, Edgewood Arsenal ILO (12, 15, 16, 17, 26)
U. S. Army, Fort Detrick (2, 3, 21, 26, 42, 43, 44)

LITERATURE REFERENCES

(1) Anon., "FDA, Medicine Library to Test Wiswesser Line Notation," **Chem. Eng. News 46,** 21-22 (1968).

(2) Barnard, A. J., Jr., C. T. Kleppinger and W. J. Wiswesser, "Retrieval of Organic Structures from Small-to-Medium Sized Collections," **J. Chem. Doc. 6,** 41-48 (1966).

(3) Barnard, A. J. Jr., and W. J. Wiswesser, "Computer-Serviced Management of Chemical Structure Information," **Lab. Management, 5,** 34-44 (October 1967).

(4) Berry, M. M., and J. W. Perry, "Notational Systems for Structural Formulas," **Chem. Eng. News 30,** 407-410 (1952).

(5) Bonnett, H. T., and D. W. Calhour, "Application of a Line-Formula Notation in an Index of Chemical Structures," **J. Chem. Doc. 2,** 2-6 (1962).

(6) Bonnett, H. T., "Chemical Notations – A Brief Review," **J. Chem. Doc. 3,** 235-242 (1963).

(7) Bowman, C. M., F. A. Landee, and M. H. Reslock "A Chemically Oriented Information Storage and Retrieval System," **J. Chem. Doc. 7,** 43-47 (1967).

(8) Bowman, C. M., F. A. Landee, N. W. Lee, and M. H. Reslock, "Computer Generation of the Wiswesser Notations of Complex Polycyclic Structures," **J. Chem. Doc. 8,** 133-138 (1968).

(9) Chemical Abstracts Service, "Wiswesser Line Notations Corresponding to Ring Index Structures," a report to NSF, ACS (1968); available from CLEARINGHOUSE for Fed. Sci. & Tech. Info., Springfield, Va. 22151 (PB 180901).

(10) Davis, J. E., and L. E. Straka, "Automatic Coding Via a Cathode Ray Tube," paper presented at the 158th National Meeting, Am. Chem. Soc., New York, September 1969.

(11) Fraction G. F., J. C. Walker, and S. J. Tauber, "Connection Tables from Wiswesser Structure Notations – A Partial Algorithm," NBS Tech. Note 432, Natl. Bur of Standards, Washington, D.C., 1968.

(12) Gelberg, A. W. Nelson, G. S. Yee, and E. A. Metcalf, "A Program for Retrieval of Organic Structure Information," **J. Chem. Doc. 2,** 7-11, (1962).

(13) Gelberg, A., "Rapid Structure Searches via Permuted Chemical Line-Notations. IV, A Reactant Index," **J. Chem. Doc. 6,** 60-61 (1966).

(14) Gelberg, A., "Information Centers in Management and Planning," Trans. of the 11th Ann. Meeting, Am. Assn. Cost Engs., Cleveland, Ohio, 1967.

(15) Gibson, G. W., C. E. Granito, D. E. Renard, and E. A. Metcalf, "The Wiswesser Line Notation: An Introduction," Tech. Memo 7-3, Edgewood Arsenal, Md., 1965. Available from the CLEARINGHOUSE, Springfield, Va. 22151 (AD 624525).

(16) Granito, C. E., J. E. Schultz, G. W. Gibson, A. Gelberg, R. J. Williams, and E. A. Metcalf, "Rapid Structure Searches via Permuted Chemical Line-Notations. III, A Computer-Produced Index," **J. Chem. Doc. 5,** 229-233 (1965).

(17) Granito, C. E., D. E. Renard, and L. A. Holly, "Use of the Wiswesser Line-Notation for Determining Duplicate Chemical Structures," **Ibid. 6,** 252-253 (1966).

(18) Haglind, J. B., H. J. Ackermann, R. E. Maizell, T. N. Manning, and B. S. Schlessinger, "Storage and Retrieval of Agricultural Screening Data," paper presented at the 156th Natl. Meeting, Am. Chem. Soc., Atlantic City, Sept. 1968.

(19) Hoffmann, E., "Use of a Modified Wiswesser Notation for Encoding of Proteins," **J. Chem. Doc. 9,** 137 (1969).

(20) Horner, J. K., "Low-Cost Storage and Retrieval of Organic Structures by Permuted Line Notations," **J. Chem. Doc. 7,** 85-88 (1967).

(21) Hyde, E., F. W. Matthews, L. H. Thomson, and W. J. Wiswesser, "Conversion of Wiswesser Notation to a Connectivity Matrix for Organic Compounds," **ibid., 7,** 200-204 (1967).

(22) Hyde, E., and L. H. Thomson, "Structure Display," **ibid., 8,** 138-146 (1968).

(23) Institute for Scientific Information, "The Index Chemicus Registry System," ISI, Philadelphia, Pa. 1968.

(24) Kulpinski, S., N. London, D. Lefkovitz, and A. Gennaro, "A Study and Implementation of Mechanical Translation from Wiswesser Line-Notation to Connection Table," Univ. of Pa., Philadelphia, 1967 (PB 177292).

(25) Lynch, M. F., "Conversion of Connection Table Descriptions of Chemical Compounds into a Form of Wiswesser Notation," **J. Chem. Doc. 8,** 130-133 (1968).

(26) Mitchell, J. P., Editor, "Proceedings of the Wiswesser Line Notation Meeting of the Army CIDS Program," EASP 400-408, Edgewood Arsenal, 1968; available from CLEARINGHOUSE for Fed. Sci. & Tech. Info. (AD 665397).

(27) Ofer, K. D., "A Computer Program to Index or Search Linear Notations," **J. Chem. Doc. 8,** 128-129 (1968).

(28) Revesz, G. S. and A. Warner, "Index Chemicus, an Important Tool for Chemical Information Processing and Retrieval," paper presented at the 156th Natl. Mtg., Am. Chem. Soc., Atlantic City, September 1968.

(29) Rice, C. N., K. D. Ofer, and R. B. Bourne, "A Pilot Study for the Input to a Chemical-Structure Retrieval System," paper presented at the 151st Natl. Mtg., Am. Chem. Soc., Chicago, September 1966.

(30) Rice, C. N., "Toward a National Systems Resource in Toxicology," **J. Chem. Doc. 9,** 181-183 (1969).

(31) Smith, E. G., "Machine Searching for Chemical Structures," **Science 131,** 142-146 (1960).

(32) Smith, E. G., "The Wiswesser Line-Formula Chemical Notation," McGraw-Hill Book Co., New York, 1968. 309 pages.

(33) Smith, E. G., "The Wiswesser Notation: State of the Art in 1969," a report to the Chemical Notation Association, September 1969.

(34) Sorter, P. F., G. E. Granito, J. C. Gilmer, A. Gelberg, and E. A. Metcalf, "Rapid Structure Searches via Permuted Chemical Line-Notations," **J. Chem. Doc. 4,** 56-60 (1964).

(35) Starke, A. C., F. R. Whaley, E. C. Carson, and W. B. Thompson, "The GAF Document Storage and Retrieval System," **Am. Doc. 19,** 173-180, (1968).

(36) Thomson, L. H., E. Hyde, and F. W. Matthews, "Organic Search and Display Using a Connectivity Matrix Derived from Wiswesser Notation," **J. Chem. Doc. 7,** 204-209 (1967).

(37) Wiswesser, W. J., "Simplified Chemical Coding for Automatic Sorting and Printing Machinery," Willson Products, Inc., Reading, Pa., 1951 (out of print).

(38) Wiswesser, W. J., "The Wiswesser Line-Formula Notation," **Chem. Eng. News 30,** 3523-3526 (1952).

(39) Wiswesser, W. J., "A Systematic Line-Formula Chemical Notation," **Aslib. Proc. 5,** 137-147 (1953).

(40) Wiswesser, W. J., "A Line-Formula Chemical Notation," Thomas Y. Crowell Co., New York, 1954. 149 pages (out of print; available from author).

(41) Wiswesser, W. J., "Literature Sources of Mammalian Toxicity Data, with Special Emphasis on Tabulating Machinery Applications," in Advances in Chemistry Series No. 16, 64-82, Amer. Chem. Soc., Washington, D.C., 1956.

(42) Wiswesser, W. J., "107 Years of Line-Formula Notations (1861-1968)," **J. Chem. Doc. 8,** 146-150 (1968).

(43) Wiswesser, W. J., and A. J. Barnaid, Jr., "The Retrieval of Chemical Structure Information," **Brit. Soc. Rheol. Bull.,** No. 2, 40-58, (1968).

(44) Wiswesser, W. J., "Computer Applications of the WLN," Tech. MS 490, Fort Detrick, Frederick, Md., January 1969.